Discrete choice and stochastic utility maximization

Koning, Ruud H.; Ridder, G.

MEMORANDUM

from

INSTITUTE OF ECONOMIC RESEARCH

FACULTY OF ECONOMICS

UNIVERSITY OF GRONINGEN

P.O. Box 800

9700 AV Groningen

Research Memorandum nr. 414

DISCRETE CHOICE AND STOCHASTIC UTILITY MAXIMIZATION

by

Ruud H. Koning
Geert Ridder

The authors would like to thank Andries and Jan Karel Lenstra for helpful comments, without implicating them for remaining errors.

March 1991

Abstract

In this paper we study the relationship between discrete or qualitative choice models and the hypothesis of stochastic utility maximization. Except for computational problems it is relatively easy to derive choice probabilities from a given random utility specification. We consider the reverse question: under which conditions is a given set of choice probabilities compatible with stochastic utility maximization? We distinguish between global compatibility, *i.e.* compatibility for all possible values of the observed utility components, and local compatibility, *i.e.* compatibility on a subset of the Euclidean space of appropriate dimension.

We give a simple derivation of the necessary and sufficient conditions for global compatibility. We also discuss local compatibility on various subsets, and we give a necessary and sufficient condition for local compatibility given a natural choice of this subset.

Keywords and phrases: Discrete choice, stochastic utility maximization, random utility models.

JEL classification: 022, 211.

1 Introduction

Consider an economic agent who must choose between I alternatives that are indexed by $i = 1, \ldots, I$. An econometrician wants to predict the choice made by the agent. An obvious procedure is to measure the utility of the alternatives to the agent, denoted by u_i, $i = 1, \ldots, I$, and to predict that he chooses the alternative that yields the highest level of utility. If utility is measurable, the econometrician can easily check whether the observed choice is consistent with utility maximization.

In practice, the econometrician has only limited knowledge of the utility attached to alternative i. We assume that this lack of knowledge can be adequately represented by letting the $I-$vector of utilities be a draw from an $I-$variate distribution. The econometrician knows the mean of this distribution. Hence we can write

$$u_i = -v_i + \varepsilon_i, \; i = 1, \ldots, I \tag{1}$$

where $-v_i$ is the known mean of u_i, and ε_i is a zero mean random variable. The joint distribution function of the $I-$vector ε is denoted by F. This joint distribution function is known to the econometrician, and is absolutely continuous (with respect to the Lebesgue measure in $\mathbb{R}^I$). The model in equation (1) is the well-known random utility model.

The fact that the utilities of the alternatives are only known up to a mean zero random variable makes the behavior of the agent harder to predict. The econometrician can only specify choice probabilities $P_i(v)$, $i = 1, \ldots, I$. If the agent chooses the alternative with the highest level of utility, *i.e.* if he prefers i over j if and only if $u_i > u_j$, then the choice probabilities take the following well-known form:

$$P_i(v) = \int_{-\infty}^{\infty} \frac{\partial F}{\partial \varepsilon_i}(\varepsilon_i - v_i + v_1, \ldots, \varepsilon_i, \ldots, \varepsilon_i - v_i + v_I) d\varepsilon_i,$$

$$i = 1, \ldots, I.$$

How can the econometrician check whether the agent maximizes his utility? Clearly, a single observed choice is not sufficient. Let us assume that the econometrician by observing choices made by a large group of agents is able to determine the choice probabilities $P_i(v)$, $i = 1, \ldots, I$ for all $v \in \mathbb{R}^I$.

Alternatively, these choice probabilities are obtained by fitting a flexible functional form to the observed choices of a finite number of agents. Necessary and sufficient conditions for the compatibility of these choice probabilities with stochastic utility maximization have been known for some time (see Daly and Zachary (1979) and McFadden (1981)). Section 2 of this paper contains a derivation of these conditions, that is much simpler than available in the literature.

As noted by Börsch-Supan (1990) it is possible that the choice probabilities only satisfy the necessary and sufficient conditions on a subset of $\mathbb{R}^I$. To be specific, Börsch-Supan assumes that the conditions hold on an $I-$dimensional interval. In general, it is not true that if the choice probabilities satisfy the necessary and sufficient conditions on a subset of v's, then choices over alternatives with these values of v are made by stochastic utility maximization. Börsch-Supan proposes a condition on the choice probabilities that ensures compatibility with utility maximization for values of v in an interval. However, as shown in section 3, his proof is not correct. A new proof is provided.

Börsch-Supan's condition is only sufficient if v is restricted to an $I-$ dimensional interval. In section 3 we show that this condition is not sufficient for compatibility on arbitrary subsets of $\mathbb{R}^I$ with positive (Lebesgue) measure. This restricts the usefulness of his condition. However, we will argue that a quest for a sufficient condition that is valid for arbitrary subsets of positive measure is unnecessary. An econometrician usually observes the choice probabilities only for a finite number of values of v, *i.e.* he observes $(P(v_t), v_t)$ for $t = 1, \ldots, T$. The question then becomes whether the choice probabilities are compatible with stochastic utility maximization for the observed values of v. Necessary and sufficient conditions for local compatibility on a finite set of v's are given in section 4. Section 5 contains an illustrative example.

2 Global Compatibility with Stochastic Utility Maximization

In this section we present a simple derivation of the necessary and sufficient conditions for the compatibility of choice probabilities $P_i(v)$, $i = 1, \ldots, I$

with stochastic utility maximization. We shall assume that v can take any value in $\mathbb{R}^I$. The necessary and sufficient conditions were first given by Daly and Zachary (1979) (see also McFadden (1981)), but these authors did not publish a proof [1]. The simple argument given here is also helpful in understanding sections 3 and 4 of the present paper.

First, we define global compatibilty.

Definition 1 *The set of choice probabilities $P_i(v)$, $i = 1, \ldots, I$ is globally compatible with stochastic utility maximization, if for all $v \in \mathbb{R}^I$ we can write for $i = 1, \ldots, I$*

$$P_i(v) = Pr(\varepsilon_j - v_j \leq \varepsilon_i - v_i; j = 1, \ldots, I, j \neq i) \tag{2}$$

with ε a stochastic $I-$vector with a non-defective and absolutely continuous (with respect to the Lebesgue measure) distribution.

It is easily seen that, if ε has an absolutely continuous distribution, then

$$\begin{aligned} &Pr(\varepsilon_j - v_j \leq \varepsilon_i - v_i; j = 1, \ldots, I, j \neq i) \\ &= \int_{-\infty}^{\infty} \frac{\partial F}{\partial \varepsilon_i}(\varepsilon_i - v_i + v_1, \ldots, \varepsilon_i, \ldots, \varepsilon_i - v_i + v_I) d\varepsilon_i, \end{aligned} \tag{3}$$

The requirement that the distribution of ε is non-defective and absolutely continuous ensures that, with probability one, the utility maximizing choice is unique.

In the sequel ι_{I-1} denotes an $(I-1)-$vector of ones. The $(I-1)-$vectors v^i and ε^i are obtained from the $I-$vectors v and ε by deleting the i-th element. Daly and Zachary (1979) state that the following conditions are necessary and sufficient for global compatibility. If a condition is subscribed by i or i and j, it holds for all $i = 1, \ldots, I$ or $i \neq j = 1, \ldots, I$, respectively.

[1] From personal communication with one of the authors we learned that they did not complete a proof. McFadden (1981) gives a proof but he deals with a situation of joint discrete/continuous choice.

Necessary and sufficient conditions for global compatibility.
For all $v \in \mathbb{R}^I$:

$$P_i(v) \geq 0, \sum_{i=1}^{I} P_i(v) = 1 \tag{C1}$$

$$P_i(v) = P_i(v + \alpha\iota_I) \text{ for all } \alpha \in \mathbb{R} \text{ (translation invariance)} \tag{C2}$$

$$\lim_{v_i \to -\infty} P_i(v) = 1, \lim_{v_j \to -\infty} P_i(v) = 0 \tag{C3}$$

$P_i(v)$ is differentiable with respect to v^i and

$$\frac{\partial^{(I-1)} P_i}{\partial v^i}(v) \geq 0 \text{ (non-negativity)} \tag{C4}$$

$$\frac{\partial P_i}{\partial v_j}(v) = \frac{\partial P_j}{\partial v_i}(v) \text{ (symmetry).} \tag{C5}$$

Theorem 1 (Daly and Zachary) *Conditions (C1)-(C5) are necessary and sufficient for global compatibility with stochastic utility maximization.*

Proof (Necessity) (C1) follows directly from the uniqueness (with probability one) of the utility maximizing choice. Translation invariance is a direct consequence of equation (3). The non-defectiveness of the distribution of ε implies (C3). The differentiability and non-negativity follows from equation (3) and the absolute continuity of the distribution of ε. Finally, we have

$$\begin{aligned}
\frac{\partial P_j}{\partial v_i}(v) &= \int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \varepsilon_i \partial \varepsilon_j}(\varepsilon_j - v_j + v_1, \ldots, \varepsilon_j, \ldots, \varepsilon_j - v_j + v_I) d\varepsilon_i \\
&= \int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \varepsilon_i \partial \varepsilon_j}(\varepsilon_i - v_i + v_1, \ldots, \varepsilon_i, \ldots, \varepsilon_i - v_i + v_I) d\varepsilon_i \\
&= \frac{\partial P_i}{\partial v_j}(v),
\end{aligned}$$

where the second equality is obtained by the change of variable $\varepsilon_i = \varepsilon_j + v_i - v_j$.
(Sufficiency) Translation invariance implies that we can write

$$P_i(v) = P_i(v - v_i\iota_I) = H_i(v^i - v_i\iota_{I-1}) \tag{4}$$

with H_i a function defined on $\mathbb{R}^{(I-1)}$. Let for $w \in \mathbb{R}^{(I-1)}$

$$h_i(w) = \frac{\partial^{(I-1)} H_i}{\partial w}(w). \tag{5}$$

Because of (C4) and equation (4) h_i exists, and is non-negative on $\mathbb{R}^{(I-1)}$. Moreover

$$P_i(v) = \int_{-\infty}^{v^i - v_i \iota_{I-1}} h_i(w) \mathrm{d}w. \tag{6}$$

Note that from (C5) for $i \neq j$ and all $v \in \mathbb{R}^I$

$$h_i(v^i - v_i \iota_{I-1}) = h_j(v^j - v_j \iota_{I-1}). \tag{7}$$

Comparison of equation (6) and equation (2) indicates that we must show that there exists a random I–vector ε with an absolutely continuous and non-defective distribution such that for $i = 1, \ldots, I$

$$w^i = \varepsilon^i - \varepsilon_i \iota_{I-1}$$

has density function h_i.

Let k be an arbitrary non-defective density function and specify the distribution function of ε by

$$F(\varepsilon) = \int_{-\infty}^{\varepsilon_1} H_1(\varepsilon^1 - \varepsilon_1 \iota_{I-1}) k(s) \mathrm{d}s.$$

The corresponding density function is, of course,

$$f(\varepsilon) = h_1(\varepsilon^1 - \varepsilon_1 \iota_{I-1}) k(\varepsilon_1).$$

Note that in the construction of F and f we started from 1 as a 'reference alternative'. A simple transformation of ε to w^1 and ε_1 shows that $\varepsilon^1 - \varepsilon_1 \iota_{I-1}$ has density function h_1. We need to show that $\varepsilon^i - \varepsilon_i \iota_{I-1}$ has density function h_i for $i = 2, \ldots, I$. Without loss of generality we choose $i = I$ (if necessary, we relabel the alternatives). Now consider the transformation from ε to

$$\begin{aligned}
\eta_1 &= \varepsilon_1 - \varepsilon_I \\
&\vdots \\
\eta_{I-1} &= \varepsilon_{I-1} - \varepsilon_I \\
\eta_I &= \varepsilon_I.
\end{aligned}$$

The corresponding density function of η is

$$g(\eta) = h_1(\eta_2 - \eta_1, \ldots, \eta_{I-1} - \eta_1, -\eta_1)k(\eta_1 + \eta_I). \tag{8}$$

It is easily seen that equation (7) implies that

$$h_1(\eta_2 - \eta_1, \ldots, \eta_{I-1} - \eta_1, \eta_I - \eta_1) = h_I(\eta_1 - \eta_I, \ldots, \eta_{I-1} - \eta_I). \tag{9}$$

Setting $\eta_I = 0$ and substituting in equation (8) gives

$$g(\eta) = h_I(\eta_1, \ldots, \eta_{I-1})k(\eta_1 + \eta_I).$$

Integrating out η_I shows that $\varepsilon^I - \varepsilon_I \iota_{I-1}$ has density function h_I.

The distribution of ε is absolutely continuous by construction (and the marginal distribution of ε is in addition non-defective). It is also non-defective, because

$$F(\varepsilon) = \int_{-\infty}^{\varepsilon_i} P_1(0, \varepsilon^1 - s\iota_{I-1})k(s)\mathrm{d}s$$

and, hence from (C3)

$$\lim_{\varepsilon_i \to -\infty} F(\varepsilon) = 0, \quad i = 2, \ldots, I$$

$$\lim_{\varepsilon^1 \to \infty} F(\varepsilon) = 1,$$

where the latter equality follows from

$$\lim_{v^1 \to \infty} P_1(0, v^1) = \lim_{v_1 \to -\infty} P_1(v - v_1\iota_I) = \lim_{v_1 \to -\infty} P_1(v) = 1.$$

□

Global compatibility implies that there exists a stochastic I-vector ε that satisfies equation (2). The proof shows that the choice of ε is not unique. The choice probabilities determine h_1, and through equation (7) also $h_2, \ldots, h_I$. In other words, they determine the distributions of $\varepsilon^1 - \varepsilon_1\iota_{I-1}, \ldots, \varepsilon^I - \varepsilon_I\iota_{I-1}$. One marginal distribution, *e.g.* the distribution of ε_1, can be chosen arbitrarily. From equation (7) it follows that any one of the h_i determines all the other h_i's. For $I = 2$ this expression reduces to

$$h_1(v_2 - v_1) = h_2(v_1 - v_2),$$

i.e., h_2 is obtained by reflection of h_1 around 0.

The proof of theorem 1 contains a useful corollary.

Corollary 1 *The choice probabilities $P_i(v)$, $i = 1, \ldots, I$ are globally compatible with stochastic utility maximization if and only if there exist density functions $h_1, \ldots, h_I$ on $\mathbb{R}^{(I-1)}$, that for all $v \in \mathbb{R}^I$, $i, j = 1, \ldots, I$ satisfy*

$$h_i(v^i - v_i\iota_{I-1}) = h_j(v^j - v_j\iota_{I-1}),\ i \neq j \tag{10}$$

and

$$P_i(v) = \int_{-\infty}^{v^i - v_i\iota_{I-1}} h_i(w)\,dw. \tag{11}$$

By a change of variables we obtain a second corollary.

Corollary 2 *The choice probabilities $P_i(v)$, $i = 1, \ldots, I$ are globally compatible with stochastic utility maximization if and only if there is a density function h_1 on $\mathbb{R}^{(I-1)}$ such that for all $v \in \mathbb{R}$ and $i = 2, \ldots, I$ we have*

$$P_1(v) = \int_{-\infty}^{v^1 - v_1\iota_{I-1}} h_1(w)\,dw, \tag{12}$$

and

$$P_i(v) = \int_{v_i - v_1}^{\infty} \int_{-\infty}^{w_{i-1} + (v_2 - v_1) - (v_i - v_1)} \cdots \int_{-\infty}^{w_{i-1} + (v_I - v_1) - (v_i - v_1)} h_1(w)\,dw_1 \cdots dw_{I-1}\,dw_{i-1}. \tag{13}$$

3 A Sufficient Condition for Local Compatibility on an Interval

In the definition of global compatibility the observed utility components v can take on any value in $\mathbb{R}^I$. In local compatibility v is restricted to a subset of $\mathbb{R}^I$. Of course, local compatibility is weaker than global compatibility.

Definition 2 *The set of choice probabilities $P_i(v)$, $i = 1,\ldots,I$ is locally compatible with stochastic utility maximization on a set $\mathcal{V} \subset \mathbb{R}^I$, if for $i = 1,\ldots,I$ and all $v \in \mathcal{V}$ we can write*

$$P_i(v) = Pr(\varepsilon_j - v_j \leq \varepsilon_i - v_i;\ j = 1,\ldots,I,\ j \neq i) \tag{14}$$

with ε a stochastic $I-$vector with a non-defective and absolutely continuous distribution.

Local compatibility was introduced by Börsch-Supan (1990), although he does not give a formal definition of the concept. Local compatibility is closer to econometric practice than global compatibility. In practice, v does not take on all values in $\mathbb{R}^I$, but we usually have a finite number of observed utility components v_t, $t = 1,\ldots,T$. We specify choice probabilities, and ask whether these choice probabilities are consistent with utility maximization on a set $\mathcal{V}$ such that $v_t \in \mathcal{V}$ for $t = 1,\ldots,T$. In Börsch-Supan's study the choice probabilites are obtained by fitting a flexible functional form, the Nested Multinomial Logit model (NMNL), to the observed v_t and the corresponding observed choices. The fitted choice probabilities satisfy all conditions (C), except the global non-negativity condition in (C4). Hence, they are not globally compatible with stochastic utility maximization.

Now choose $a, b \in \mathbb{R}^I$ such that $a \leq v_t \leq b$ for $t = 1,\ldots,T$. We can ask under which conditions the fitted choice probabilities are locally compatible with stochastic utility maximization on $\mathcal{V} = [a, b]$. The following theorem gives a sufficient condition.

Theorem 2 (Börsch-Supan) *Let $P_i(v)$, $i = 1,\ldots,I$ be a set of choice probabilities that satisfy the conditions (C1)-(C5), except that for some $v \in \mathbb{R}^I$ the non-negativity condition in (C4) is violated. A sufficient condition for local compatibility of these choice probabilities on an interval $\mathcal{V} = [a, b]$ is that for all $a \leq v \leq b$, we have*

$$\frac{\partial^k P_1}{\partial v^{1kl}}(v) \geq 0 \tag{15}$$

with v^{1kl} a $k-$dimensional subvector of the $(I-1)-$vector v^1, and $l = 1,\ldots,\binom{I-1}{k}$.

Note that equation (15) only refers to P_1. By symmetry we have that equation (15) also holds for $i = 2, \ldots, I$. We can weaken the conditions somewhat: from the proof it follows that symmetry only has to hold for $a \leq v \leq b$. Translation invariance implies that we can replace the interval $[a, b]$ by

$$D = \left\{ v \in \mathbb{R}^I \mid \exists c \in \mathbb{R} \text{ such that } a \leq v - c\iota_I \leq b \right\}$$

Börsch-Supan's proof of this theorem is not correct. His proof starts from the representation in equation (11), with h_1 defined in equation (5). Note that h_1 is not a density function because for some $v \in \mathbb{R}^I$ we have that $h_1(v^1 - v_1\iota_{I-1}) < 0$. Börsch-Supan proposes a density function h_1^* such that for all $a \leq v \leq b$

$$P_1(v) = \int_{-\infty}^{v^1 - v_1\iota_{I-1}} h_1^*(w)\mathrm{d}w \tag{16}$$

In equation (16), $h_1^*(w) \geq 0$ for all $w \in \mathbb{R}^{(I-1)}$. However, his suggestion for h_1^* does not satisfy equation (16).

To see this, first note that if $a \leq v \leq b$, then[2]

$$a^1 - b_1\iota_{I-1} \leq v^1 - v_1\iota_{I-1} \leq b^1 - a_1\iota_{I-1} \tag{17}$$

In the sequel we consider the case $I = 3$.

In equation (24), p. 382, Börsch-Supan suggests to define h_1^* by

$$h_1^*(w) = \begin{cases} h_1(w) & w_1 > a_2 - b_1, w_2 > a_3 - b_1 \\ \dfrac{\partial P_1}{\partial w_2}(w) & w_1 = a_2 - b_1, w_2 > a_3 - b_1 \\ \dfrac{\partial P_1}{\partial w_1}(w) & w_1 > a_2 - b_1, w_2 = a_3 - b_1 \\ P_1(w) & w_1 = a_2 - b_1, w_2 = a_3 - b_1 \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

[2]The interval $[c, d]$ on page 382 of Börsch-Supan's paper coincides with (17), because $a_i - b_1 = \min(a_i - b_1, b_i - a_1)$ and $b_i - a_1 = \max(a_i - b_1, b_i - a_1)$.

Because h_1^* is a density function with respect to the Lebesgue measure, an equivalent density function is given by

$$h_1^{**}(w) = \begin{cases} h_1(w) & w_1 > a_2 - b_1, w_2 > a_3 - b_1 \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

The density functions in equations (18) and (19) are equivalent because they differ on a set of Lebesgue measure 0.

It is immediately clear that for $a \leq v \leq b$

$$P_1(v) > \int_{-\infty}^{v^1 - v_1 \iota_{I-1}} h_1^{**}(w) \mathrm{d}w$$

because

$$P_1(b_1, a_2, a_3) = \int_{-\infty}^{a^1 - b_1 \iota_2} h_1(w) \mathrm{d}w > 0$$

We conclude that Börsch-Supan's construction does not satisfy equation (16). The problem is that he changes the value of a density function on a set of Lebesgue measure 0.

For a correct proof we must find a density function h_1^* that satisfies equation (16). If the density function of the $I-$vector ε is chosen as

$$f^*(\varepsilon) = h_1^*(\varepsilon^1 - \varepsilon_1 \iota_{I-1}) k(\varepsilon_1) \tag{20}$$

with k an arbitrary density function, then $P_1(v)$ can be expressed as in equation (11). For all $v \in \mathcal{V}$ and $i = 2, \ldots, I$ the density function h_1^* must also satisfy

$$\begin{aligned} &P_i(v) \\ &\quad = Pr(\varepsilon_2 - \varepsilon_1 \leq \varepsilon_i - \varepsilon_1 + (v_2 - v_1) - (v_i - v_1), \ldots, (\varepsilon_i - \varepsilon_1) \geq (v_i - v_1), \\ &\qquad \ldots, \varepsilon_I - \varepsilon_1 \leq \varepsilon_i - \varepsilon_1 + (v_I - v_1) - (v_i - v_1)) \\ &\quad = \int_{v_i - v_1}^{\infty} \int_{-\infty}^{w_{i-1} + (v_2 - v_1) - (v_i - v_1)} \cdots \int_{-\infty}^{w_{i-1} + (v_I - v_1) - (v_i - v_1)} \\ &\qquad h_1^*(w) \mathrm{d}w_1 \cdots \mathrm{d}w_{I-1} \mathrm{d}w_{i-1} \end{aligned} \tag{21}$$

(*cf.* corollary 2). Hence, if we can find a density function h_1^* such that for all $v \in \mathcal{V}$ the choice probabilities $P_1(v)$ and $P_i(v)$, $i = 2, \ldots, I$ can be expressed as in equations (16) and (21), then theorem 2 is proved. Note that, if equation (7) holds, *i.e.* if the choice probabilities satisfy the symmetry condition, then by a change of variables the choice probabilities $P_i(v)$, $i = 2, \ldots, I$ can for all $v \in \mathbb{R}^I$ be expressed as

$$P_i(v) = \int_{v_i - v_1}^{\infty} \int_{-\infty}^{w_{i-1}+(v_2-v_1)-(v_i-v_1)} \cdots \int_{-\infty}^{w_{i-1}+(v_I-v_1)-(v_i-v_1)} h_1(w) \mathrm{d}w_1 \cdots \mathrm{d}w_{I-1} \mathrm{d}w_{i-1} \qquad (22)$$

However, h_1 is not a density function so that corollary 2 does not hold.

Proof of theorem 2 We prove the theorem for $I = 3$. The proofs for $I = 4, 5, \ldots$ are notationally involved, but are completely analogous. From figure 1 we see

$$\begin{aligned} P_1(v) &= \int_{-\infty}^{v_2-v_1} \int_{-\infty}^{v_3-v_1} h_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1 \\ &= \int_{-\infty}^{a_2-b_1} \int_{-\infty}^{a_3-b_1} h_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1 + \int_{a_2-b_1}^{v_2-v_1} \int_{-\infty}^{a_3-b_1} h_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1 \\ &+ \int_{-\infty}^{a_2-b_1} \int_{a_3-b_1}^{v_3-v_1} h_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1 + \int_{a_2-b_1}^{v_2-v_1} \int_{a_3-b_1}^{v_3-v_1} h_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1 \end{aligned} \qquad (23)$$

We consider the four terms of equation (23) in turn. Because

$$\int_{-\infty}^{a_2-b_1} \int_{-\infty}^{a_3-b_1} h_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1 = P_1(0, a_2 - b_1, a_3 - b_1) \geq 0,$$

there is a density function g_1 with

$$\int_{-\infty}^{a_2-b_1} \int_{-\infty}^{a_3-b_1} h_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1 = \int_{-\infty}^{a_2-b_1} \int_{-\infty}^{a_3-b_1} g_1(w_1, w_2) \mathrm{d}w_2 \mathrm{d}w_1.$$

For the second term let f_2 be an arbitrary non-negative function with

$$\int_{-\infty}^{a_3-b_1} f_2(w_2) \mathrm{d}w_2 = 1. \qquad (24)$$

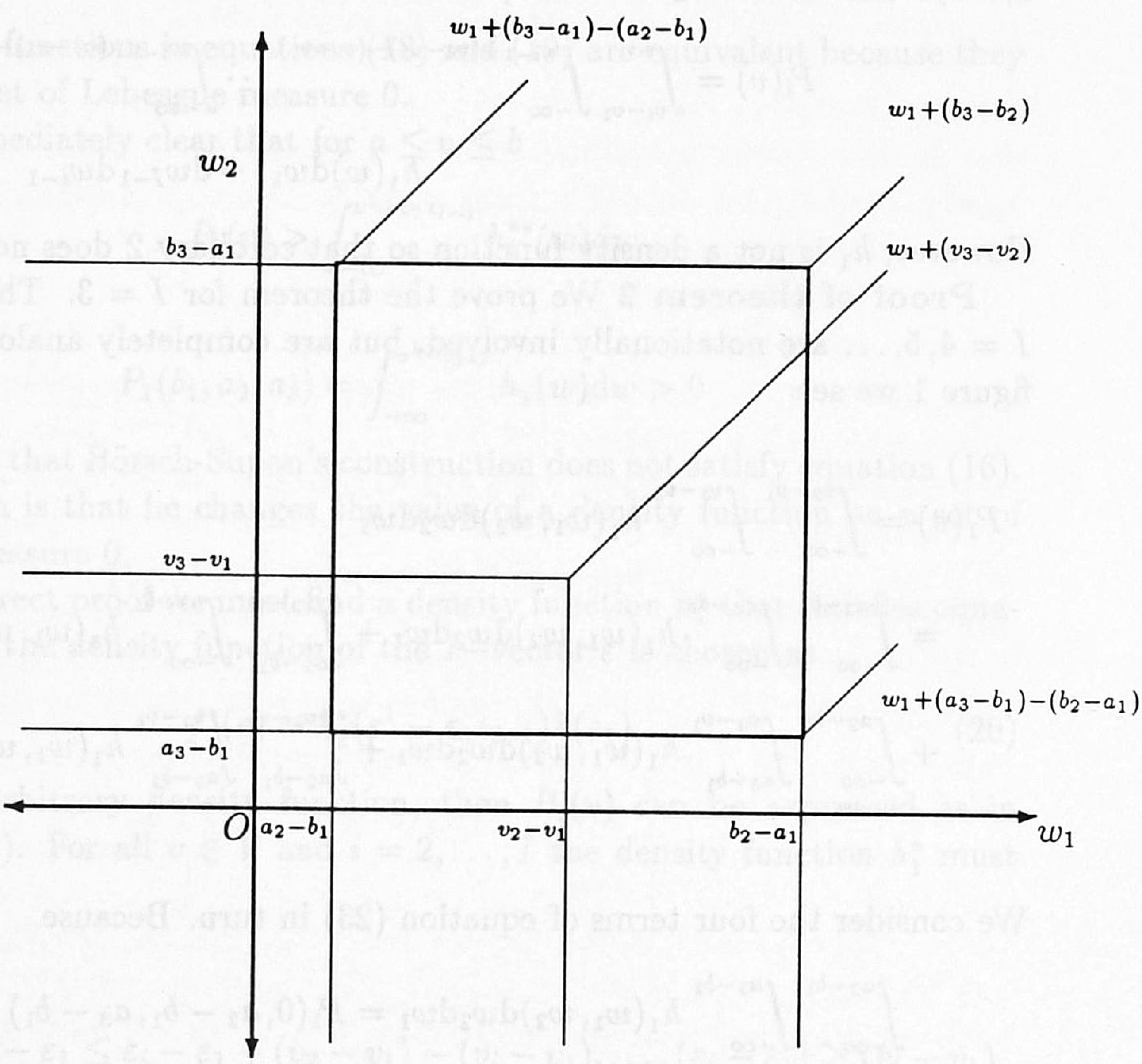

Figure 1: Integration regions for choice probabilities, $I = 3$

We have

$$\int_{a_2-b_1}^{v_2-v_1}\int_{-\infty}^{a_3-b_1} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$= \int_{a_2-b_1}^{v_2-v_1}\int_{-\infty}^{a_3-b_1}\left\{f_2(w_2)\int_{-\infty}^{a_3-b_1} h_1(w_1,s)\mathrm{d}s\right\}\mathrm{d}w_2\mathrm{d}w_1. \qquad (25)$$

Condition (15) in theorem 2 implies that for $a_2 - b_1 \leq w_1 \leq b_2 - a_1$,

$$\frac{\partial P_1}{\partial v_2}(0, w_1, a_3 - b_1) = \int_{-\infty}^{a_3-b_1} h_1(w_1,s)\mathrm{d}s \geq 0.$$

Hence, the integrand in equation (25) is non-negative. The third term is rewritten analogously. Let f_1 be an arbitrary non-negative function with

$$\int_{-\infty}^{a_2-b_1} f_1(w_1)\mathrm{d}w_1 = 1.$$

Now we can define h_1^* for $w_1 \leq b_2 - a_1$, $w_2 \leq b_3 - a_1$ as

$$h_1^*(w) = \begin{cases} h_1(w) & a_2 - b_1 \leq w_1 \leq b_2 - a_1, \\ & a_3 - b_1 \leq w_2 \leq b_3 - a_1 \\ g_1(w) & w_1 < a_2 - b_1, w_2 < a_3 - b_1 \\ f_2(w_2)\int_{-\infty}^{a_3-b_1} h_1(w_1,s)\mathrm{d}s & a_2 - b_1 \leq w_1 \leq b_2 - a_1, w_2 < a_3 - b_1 \\ f_1(w_1)\int_{-\infty}^{a_2-b_1} h_1(s,w_2)\mathrm{d}s & w_1 < a_2 - b_1, a_2 - b_1 \leq w_2 \leq b_3 - a_1. \end{cases} \qquad (26)$$

From equation (22) it follows that

$$P_2(v) = \int_{v_2-v_1}^{\infty}\int_{-\infty}^{w_1+(v_3-v_2)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$= \int_{b_2-a_1}^{\infty}\int_{-\infty}^{w_1+(a_3-b_1)-(b_2-a_1)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$+\int_{v_2-v_1}^{b_2-a_1}\int_{-\infty}^{a_3-b_1} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$+\int_{b_2-a_1}^{\infty}\int_{w_1+(a_3-b_1)-(b_2-a_1)}^{w_1+(v_3-v_2)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$+\int_{v_2-v_1}^{b_2-a_1}\int_{a_3-b_1}^{w_1+(v_3-v_2)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1 \tag{27}$$

Again we consider the four terms on the right-hand side of equation (27) in turn. Because

$$\int_{b_2-a_1}^{\infty}\int_{-\infty}^{w_1+(a_3-b_1)-(b_2-a_1)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1 = P_2(0,b_2-a_1,a_3-b_1)\geq 0$$

there is a density function g_2 with

$$\int_{b_2-a_1}^{\infty}\int_{-\infty}^{w_1+(a_3-b_1)-(b_2-a_1)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$=\int_{b_2-a_1}^{\infty}\int_{-\infty}^{w_1+(a_3-b_1)-(b_2-a_1)} g_2(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1.$$

The second term of equation (27) can be rewritten as

$$\int_{v_2-v_1}^{b_2-a_1}\int_{-\infty}^{a_3-b_1} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$=\int_{v_2-v_1}^{\infty}\int_{-\infty}^{a_3-b_1}\left\{f_2(w_2)\int_{-\infty}^{a_3-b_1} h_1(w_1,s)\mathrm{d}s\right\}\mathrm{d}w_2\mathrm{d}w_1 \tag{28}$$

with f_2 as in equation (24). The integrands in the equations (28) and (25) are identical. Hence, the integrand of equation (28) is non-negative. To deal with the third term of equation (27), let f_3 be a non-negative function with

$$\int_{b_2-a_1}^{\infty} f_3(w_1)\mathrm{d}w_1 = 1.$$

Hence,

$$\int_{b_2-a_1}^{\infty}\int_{w_1+(a_3-b_1)-(b_2-a_1)}^{w_1+(v_3-v_2)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$= \int_{b_2-a_1}^{\infty} \int_{(a_3-b_1)-(b_2-a_1)}^{v_3-v_2} h_1(w_1, w_1+w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$= \int_{b_2-a_1}^{\infty} \int_{(a_3-b_1)-(b_2-a_1)}^{v_3-v_2} \left\{ f_3(w_1) \int_{b_2-a_1}^{\infty} h_1(s, w_2+s)\mathrm{d}s \right\} \mathrm{d}w_2\mathrm{d}w_1$$

$$= \int_{b_2-a_1}^{\infty} \int_{w_1+(a_3-b_1)-(b_2-a_1)}^{w_1+(v_3-v_2)} \left\{ f_3(w_1) \int_{b_2-a_1}^{\infty} h_1(s, w_2-w_1+s)\mathrm{d}s \right\} \mathrm{d}w_2\mathrm{d}w_1. \tag{29}$$

By condition (15) of theorem 2 we have for $a_3 - b_1 \leq w_2 \leq b_3 - a_1$

$$\frac{\partial P_2}{\partial v_3}(0, b_2-a_1, w_2) = \int_{b_2-a_1}^{\infty} h_1(s, s+w_2-(b_2-a_1))\mathrm{d}s \geq 0,$$

or

$$\int_{b_2-a_1}^{\infty} h_1(s, s+r)\mathrm{d}s \geq 0$$

for $(a_3-b_1)-(b_2-a_1) \leq r \leq b_3-b_2$. Hence the integrand in equation (29) is non-negative. The integrand of the final term in equation (27) is non-negative by equation (15).

We can now extend the definition of h_1^*:

$$h_1^*(w) = \begin{cases} g_2(w) & w_1 > b_2-a_1, w_2 < w_1 + (a_3-b_1) - (b_2-a_1) \\ f_3(w_1) \int_{b_2-a_1}^{\infty} h_1(s, w_2-w_1+s)\mathrm{d}s & \\ \qquad w_1 > b_2-a_1, w_1 + (a_3-b_1) - (b_2-a_1) \leq w_2 \leq w_1 + (b_3-b_2). & \end{cases} \tag{30}$$

By equations (25) and (28) the definition on h_1^* on $a_2 - b_1 \leq w_1 \leq b_2 - a_1, w_2 < a_3 - b_1$ in equation (26) can be used for $P_2(v)$ too.

Finally, we consider $P_3(v)$. From figure 1 we obtain

$$P_3(v) = \int_{v_3-v_1}^{\infty} \int_{-\infty}^{w_2+(v_2-v_3)} h_1(w_1, w_2)\mathrm{d}w_1\mathrm{d}w_2$$

$$= \int_{b_3-a_1}^{\infty} \int_{-\infty}^{w_2+(a_2-b_1)-(b_3-a_1)} h_1(w_1, w_2)\mathrm{d}w_1\mathrm{d}w_2$$

$$+\int_{v_3-v_1}^{b_3-a_1}\int_{-\infty}^{a_2-b_1} h_1(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2$$

$$+\int_{b_2-a_1}^{\infty}\int_{w_1+(v_3-v_2)}^{w_1+(b_3-b_2)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$+\int_{b_3-a_1}^{\infty}\int_{w_2+(a_2-b_1)-(b_3-a_1)}^{w_2+(b_2-b_3)} h_1(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2$$

$$+\int_{a_2-b_1}^{v_2-v_1}\int_{v_3-v_1}^{b_3-a_1} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

$$+\int_{v_2-v_1}^{b_3-a_1}\int_{w_1+(v_3-v_2)}^{b_3-a_1} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1 \tag{31}$$

We implicitly assume that the line $w_2 = w_1 + (v_3 - v_2)$ lies below the line $w_2 = w_1 + (b_3 - b_2)$. However, the argument which follows can easily be adapted to handle the other case as well. The first term on the right-hand side of equation (31) is equal to $P_3(0, a_2 - b_1, b_3 - a_1)$, and we can find a density function g_3 such that

$$\int_{b_3-a_1}^{\infty}\int_{-\infty}^{w_2+(a_2-b_1)-(b_3-a_1)} g_3(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2 = P_3(0,a_2-b_1,b_3-a_1)\geq 0.$$

For the second term we easily check that h_1 can be replaced by h_1^* as defined in equation (26). For the third term we replace h_1 by h_1^* as defined in equation (30):

$$\int_{b_2-a_1}^{\infty}\int_{w_1+(v_3-v_2)}^{w_1+(b_3-b_2)}\left\{f_3(w_1)\int_{b_2-a_1}^{\infty} h_1(s, w_2-w_1+s)\mathrm{d}s\right\}\mathrm{d}w_2\mathrm{d}w_1$$

$$=\int_{b_2-a_1}^{\infty}\int_{v_3-v_2}^{b_3-b_2} h_1(w_1,w_2+w_1)\mathrm{d}w_2\mathrm{d}w_1$$

$$=\int_{b_2-a_1}^{\infty}\int_{w_1+(v_3-v_2)}^{w_1+(b_3-b_2)} h_1(w_1,w_2)\mathrm{d}w_2\mathrm{d}w_1$$

Let f_4 be an arbitrary non-negative function with

$$\int_{b_3-a_1}^{\infty} f_4(w_2)\mathrm{d}w_2 = 1.$$

Then we can replace the integrand in the fourth term of equation (31) by

$$f_4(w_2)\int_{b_3-a_1}^{\infty} h_1(w_1 - w_2 + s, s)\mathrm{d}s$$

which is non-negative on the integration region. Hence, we can complete the definition of h_1^* by

$$h_1^*(w) = \begin{cases} g_3(w) & w_2 > b_3 - a_1, w_1 < w_2 + (a_2 - b_1) - (b_3 - a_1) \\ f_4(w_2)\int_{b_3-a_1}^{\infty} h_1(w_1 - w_2 + s, s)\mathrm{d}s & \\ \quad w_2 > b_3 - a_1, w_2 + (a_2 - b_1) - (b_3 - a_1) \leq w_1 \leq w_2 + (b_2 - b_3) & \end{cases}$$

Note that h_1^* integrates to 1, because $P_1 + P_2 + P_3 = 1$. Hence, h_1^* is a density function. □

Theorem 2 gives a sufficient condition for local compatibility with stochastic utility maximization on an *interval.* It is natural to ask whether condition (15) is also sufficient for local compatibility on arbitrary open sets $\mathcal{V}$. The following simple example shows that theorem 2 does not hold for arbitrary open sets $\mathcal{V}$.

We consider choice between two alternatives, *i.e.* $I = 2$. The choice probabilities $P_1(v)$ and $P_2(v)$ can be expressed as

$$P_1(v) = \int_{-\infty}^{v_2 - v_1} h_1(w)\mathrm{d}w,$$

$$P_2(v) = \int_{v_2 - v_1}^{\infty} h_1(w)\mathrm{d}w,$$

with $h_1(w)$ defined in equation (5). Let h_1 and H_1, defined in equation (4), be as in figure 2.

Note that h_1 is non-negative on $\mathcal{V}_1 = (-\infty, w_0] \cup [w_1, \infty)$. However, it is not possible to find a density function h_1^* that coincides with h_1 on $\mathcal{V}_1$ and also satisfies

$$P_1(v) = \int_{-\infty}^{v_2 - v_1} h_1^*(w)\mathrm{d}w,$$

$$P_2(v) = \int_{v_2 - v_1}^{\infty} h_1^*(w)\mathrm{d}w,$$

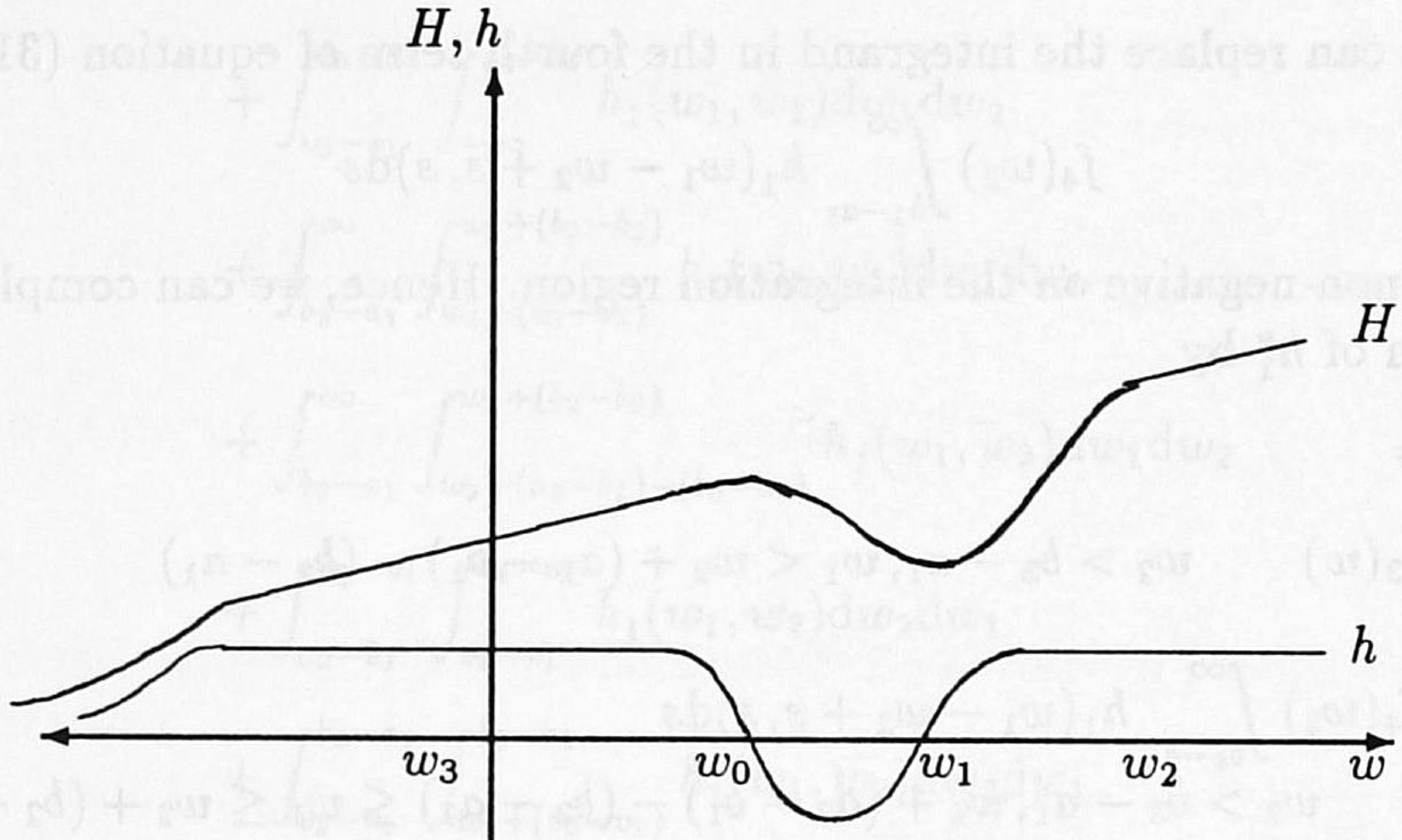

Figure 2: Integration regions for choice probabilities, $I = 3$

The obvious problem is that

$$P_1(0, w_1) < P_1(0, w_0)$$

and this implies that h_1^* has to be negative for some values of w. We conclude that although the sufficient condition of theorem 2 holds, the choice probabilities are not compatible with stochastic utility maximization on $\mathcal{V}_1$.

In this simple example it is not difficult to find sets $\mathcal{V}$ on which the choice probabilities are locally compatible. From figure 2 we see that local compatibility holds on either $\mathcal{V}_2 = (-\infty, w_3] \cup [w_1, \infty)$ or $\mathcal{V}_3 = (-\infty, w_0] \cup [w_2, \infty)$. Actually, local compatibility holds on sets $\mathcal{V}_4 = (-\infty, w'] \cup [w'', \infty)$ if and only if

$$P_1(0, w') \leq P_1(0, w'').$$

This example clearly shows the limitations of theorem 2 in checking local compatibility. Assume that we have observed utilities v_t, $t = 1, \ldots, T$. If $v_{2t} - v_{1t} \in \mathcal{V}_1$ for $t = 1, \ldots, T$ we would conclude from condition (15) of theorem 2 that the choice probabilities are compatible with stochastic utility maximization. Of course, this conclusion is incorrect. If $v_{2t} - v_{1t} \in \mathcal{V}_2$ or

$v_{2t} - v_{1t} \in \mathcal{V}_3$ for all t we correctly conclude from condition (15) that the choice probabilities are locally compatible. Note that theorem 2 in this case concludes that the choice probabilities are *not* compatible with stochastic utility maximization because the interval containing $\mathcal{V}_2$ or $\mathcal{V}_3$ is $(-\infty, \infty)$ and h_1 is negative on (w_0, w_1). From the choice of $\mathcal{V}_4$ we see that even if $v_{2t} - v_{1t} \notin \mathcal{V}_1$ or $\mathcal{V}_2$ for some t, there still is hope that the choice probabilities are locally compatible. A necessary and sufficient condition in the case $I = 2$ is that for all $s, t = 1, \ldots, T$ we have

$$v_{2t} - v_{1t} \geq v_{2s} - v_{1s} \Rightarrow P_1(v_{1t}, v_{2t}) \geq P_1(v_{1s}, v_{2s}). \tag{32}$$

Note that condition (32) does not preclude that $w_0 \leq v_{2t} - v_{1t} \leq w_1$ for some t. In other words, condition (15) is also not necessary for local compatibility of the observed choice probabilities $P_i(v_t), t = 1, \ldots, T$ with stochastic utility maximization. We conclude that if one has observed utilities v_t, $t = 1, \ldots, T$, it does not make sense to investigate local compatibility by finding a and b such that $a \leq v_{2t} - v_{1t} \leq b$ and checking whether condition (15) of theorem 2 holds for $\mathcal{V} = [a, b]$.

In the next section we propose an alternative approach that can be seen as a generalization of condition (32) for $I \geq 3$. As a bonus we obtain necessary and sufficient conditions for local compatibility with stochastic utility maximization.

4 A Necessary and Sufficient Condition for Local Compatibility on a Finite Set

In practice, an econometrician has a finite sample $\mathcal{V} = \{v_1, \ldots, v_T\}$ of observed utility components. If for v_t he observes a large number of choices made by distinct agents, he can determine the corresponding choice probabilities $P_i(v_t)$, $i = 1, \ldots, I$. If the number of observed choices per t is small, he can use either a local averaging method, *e.g.* a kernel estimate, or a flexible functional form, *e.g.* the NML model, to estimate $P_i(v_t)$, $i = 1, \ldots, I$, $t = 1, \ldots, T$. How can he decide whether these (estimated) choice probabilities are compatible with stochastic utility maximization, *i.e.* how can he decide whether the choice probabilities are locally compatible with stochastic utility maximization on $\mathcal{V}$? In this section we give necessary and sufficient conditions for local compatibility on a set that consists of a finite number of distinct points in $\mathbb{R}^I$.

The derivation of these conditions is facilitated by some additional notation. From corollary 2 we see that every choice probability can be written as

$$P_i(v) = \int_{B_i(v)} h_1(w)dw,$$

with

$$B_1(v) = \left\{w \in \mathbb{R}^{(I-1)} \mid w \leq v^1 - v_1 \iota_{I-1}\right\}$$

$$B_i(v_t) = \left\{w \in \mathbb{R}^{(I-1)} \mid w_{i-1} \geq v_i - v_1, w_j - w_{i-1} \leq (v_j - v_1)\right.$$
$$\left. -(v_i - v_1),\ j = 1, \ldots, I-1,\ j \neq i-1\right\},\ i = 2, \ldots I.$$

In this formulation, alternative 1 is chosen as the reference alternative. The choice of the reference alternative is arbitrary. Each observation v_t induces a partition of $\mathbb{R}^{(I-1)}$ into I disjoint sets $B_i(v_t)$:

$$\bigcup_{i=1}^{I} B_i(v_t) = \mathbb{R}^{(I-1)}$$

$$B_i(v_t) \cap B_j(v_t) = \emptyset,\ i \neq j.$$

For a given sample v_t, $t = 1, \ldots, T$, we define the sets C as the intersections:

$$C_{i_1 i_2 \cdots i_T} \equiv B_{i_1}(v_1) \cap B_{i_2}(v_2) \cap \ldots \cap B_{i_T}(v_T) \subset \mathbb{R}^{(I-1)} \tag{33}$$

for all $(i_1, i_2, \ldots, i_T)$ in the index set

$$J = \{(i_1, i_2, \ldots, i_T) \mid i_s = 1, \ldots, I,\ s = 1, \ldots, T\}.$$

This notation will be clarified in the example given in figure 3 below. There we have for example $C_{213} = B_2(v_1) \cap B_1(v_2) \cap B_3(v_3)$. The sets $C_{i_1 i_2 \cdots i_T}$ will be empty for many combinations of $i_1, i_2, \ldots, i_T$. For example, $B_1(v_1) \subset B_1(v_2)$ implies that $B_1(v_1) \cap B_i(v_2) = \emptyset$ for $i = 2, \ldots, I$. Furthermore, note that each set $B_i(v_t)$ can be written as the union of various sets C:

$$B_i(v_t) = \bigcup_{i(v_t)} C_{i_1 i_2 \cdots i_T},$$

where the index set $J_i(v_t)$ is given by

$$J_i(v_t) = \{(i_1, i_2, \ldots, i_T) \mid i_t = i,\ i_s = 1, \ldots, I,\ s = 1, \ldots, T,\ s \neq t\}.$$

From now on, we restrict ourselves to those sets C which are not empty, *i.e.* those belonging to

$$\mathcal{C} \equiv \{C_{i_1 i_2 \cdots i_T} \mid C_{i_1 i_2 \cdots i_T} \neq \emptyset,\ i_t = 1, \ldots, I,\ t = 1, \ldots, T\}.$$

The collection $\mathcal{C}$ is a partition of $\mathbb{R}^{(I-1)}$: the sets in $\mathcal{C}$ are disjoint and the union of all sets in $\mathcal{C}$ is $\mathbb{R}^{(I-1)}$. Using this, we can rewrite each observed choice probability as

$$P_i(v_t) = \int_{B_i(v_t)} h_1(w)\mathrm{d}w = \sum_{(i_1, i_2, \ldots, i_T) \in J_i^*(v_t)} \int_{C_{i_1 i_2 \cdots i_T}} h_1(w)\mathrm{d}w, \tag{34}$$

where $J_i^*(v_t) = \{(i_1, i_2, \ldots, i_T) \in J_i(v_t) \mid C_{i_1 i_2 \cdots i_T} \in \mathcal{C}\}$. As a last bit of notation, define

$$A_{i_1 i_2 \cdots i_T} = \int_{C_{i_1 i_2 \cdots i_T}} h_1(w)\mathrm{d}w. \tag{35}$$

Now we are in a position to give a necessary and sufficient condition for local compatibility of the observed choice probabilities on $\mathcal{V}$:

Theorem 3 *The choice probabilities $P_i(v)$, $i = 1, \ldots, I$ are locally compatible with stochastic utility maximization on $\mathcal{V} = \{v_1, \ldots, v_T\}$ if and only if*

$$P_i(v_t) = \sum_{(i_1, i_2, \ldots, i_T) \in J_i^*(v_t)} A_{i_1 i_2 \cdots i_T} \tag{36}$$

($t = 1, \ldots, T$, $i_t = 1, \ldots, I$) has a non-negative solution (for the $A_{i_1 i_2 \cdots i_T}$'s).

Proof (Necessity) It is clear form equation (34) and equation (35) that all A's will be non-negative if a non-negative generating density function exists. (Sufficiency) Suppose the set of equations (36) has a non-negative solution. It follows from equation (35) that we can construct a non-negative density $h_1(w)$ which generates the observed choice probabilities. Let for

$(i_1, i_2, \ldots, i_T) \in J^*(v_t) = \cup_{i=1}^{I} J_i^*(v_t)$ an arbitrary non-negative function $g_{i_1 i_2 \cdots i_T}$ be, such that[3]:

$$\int_{C_{i_1 i_2 \cdots i_T}} g_{i_1 i_2 \cdots i_T}(w) \mathrm{d}w = A_{i_1 i_2 \cdots i_T}.$$

We define h_1^* by

$$h_1^*(w) = g_{i_1 i_2 \cdots i_T}(w), \;\; w \in C_{i_1 i_2 \cdots i_T}, \; (i_1, i_2, \ldots, i_T) \in J^*(v_t).$$

It is clear that h_1^* is non-negative and that for $i = 1, \ldots, I$, $t = 1, \ldots, T$:

$$P_i(v_t) = \int_{B_i(v_t)} h_1^*(w) \mathrm{d}w.$$

As noted in the proof of theorem 2, this implies that there exists a stochastic I-vector ε, such that the choice probabilities can be written as in equation (14). □

In the theorem we restrict the index set to $J_i^*(v_t)$. We could also use the index set $J_i(v_t)$, but this would make the number of variables unnecessarily large.

A disadvantage of the condition in theorem 3 is that the number of equations will be large: there will be $T \times (I-1)$ equations. Moreover, determination of the index set $J_i^*(v_t)$ is a tedious task.

As a corollary of theorem 3, we can give a necessary condition on the choice probabilities which can be easily checked.

Corollary 3 *If the observed choice probabilities are compatible with stochastic utility maximization, we have for each pair (t, t'): that*

$$B_i(v_t) \subseteq B_i(v_{t'}) \Rightarrow P_i(v_t) \leq P_i(v_{t'}). \tag{37}$$

---

[3] One possible choice is

$$g_{i_1 i_2 \cdots i_T}(w) = \begin{cases} = \dfrac{A_{i_1 i_2 \cdots i_T}}{\mu(D_{i_1 i_2 \cdots i_T})} & w \in D_{i_1 i_2 \cdots i_T} \subseteq C_{i_1 i_2 \cdots i_T} \\ = 0 & w \in C_{i_1 i_2 \cdots i_T} \setminus D_{i_1 i_2 \cdots i_T}, \end{cases}$$

with $D_{i_1 i_2 \cdots i_T}$ a bounded subset of $C_{i_1 i_2 \cdots i_T}$ and $\mu(D_{i_1 i_2 \cdots i_T}) = \int_{D_{i_1 i_2 \cdots i_T}} \mathrm{d}w$.

If $I = 3$, it is easily seen that each pair of points (t, t') yields two restrictions of the form (37) (see also figure 3). In this case, there will be $T \times (T-1)$ restrictions like (37) in total.

If condition (37) is violated for some pair (t, t'), one can conclude that the choice probabilities are not locally compatible with stochastic utility maximization. If $I = 2$, the condition of corollary 3 garantuees that the distribution function of $\varepsilon_2 - \varepsilon_1$ is non-decreasing as, of course, it should be, in $v_{2t} - v_{1t}$, $t = 1, \ldots, T$. For $I = 2$ the condition of corollary 3 is also sufficient for local compatibility with stochastic utility maximization. This result, however, does not generalize to more alternatives, as the following example will illustrate[4].

Consider the case $I = 3$, $T = 3$. Let the first alternative be the reference alternative, and let $v_t^1 - v_{1t}\iota_2$ be denoted by w_t. The points w_1, w_2, w_3 and the integration region of the corresponding choice probabilities are given in figure 3. The integration regions are decomposed in sets $C_{i_1 i_2 i_3}$ defined in equation (33). The corrsponding vectors with choice probabilities are $P(v_1) = (0,0,1)'$, $P(v_2) = (0,1,0)'$ and $P(v_3) = (1,0,0)'$, which are all non-negative and sum to one for each observation.

One can easily check that the observed choice probabilities satisfy the condition of corollary 3:

$$\begin{array}{lcll}
B_1(v_1) \subseteq B_1(v_2) & \Rightarrow & P_1(v_1) \leq P_1(v_2), & (0 \leq 0) \\
B_3(v_2) \subseteq B_3(v_1) & \Rightarrow & P_3(v_2) \leq P_3(v_1), & (0 \leq 1) \\
B_1(v_1) \subseteq B_1(v_3) & \Rightarrow & P_1(v_1) \leq P_1(v_3), & (0 \leq 1) \\
B_2(v_3) \subseteq B_2(v_1) & \Rightarrow & P_2(v_3) \leq P_2(v_1), & (0 \leq 0) \\
B_2(v_3) \subseteq B_2(v_2) & \Rightarrow & P_2(v_3) \leq P_2(v_2), & (0 \leq 1) \\
B_3(v_2) \subseteq B_3(v_3) & \Rightarrow & P_3(v_2) \leq P_3(v_3), & (0 \leq 0)
\end{array}$$

However, the observed choice probabilities do not satisfy the condition of theorem 3. The set of equations

---

[4]We owe this example to Jan Karel Lenstra

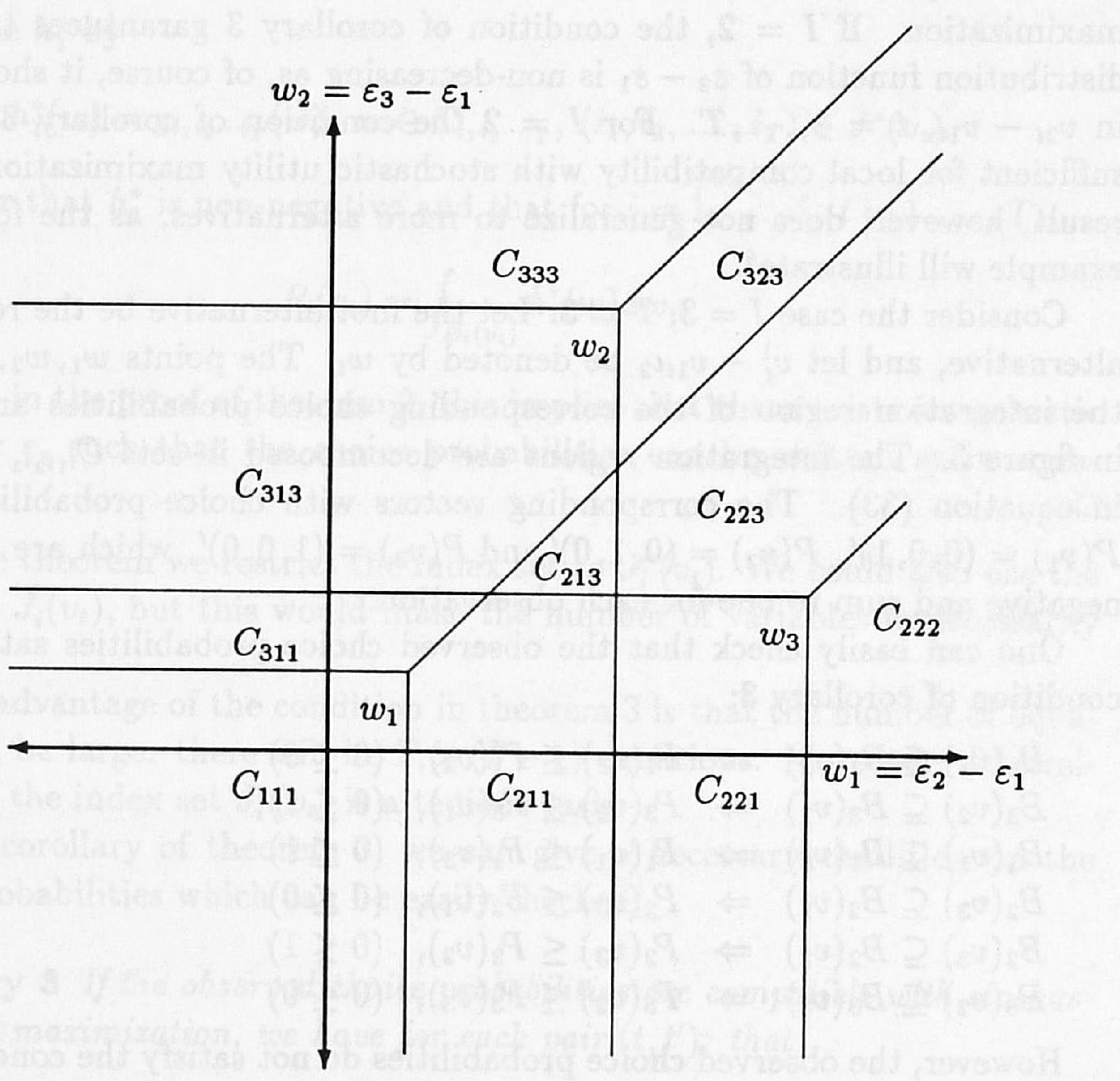

Figure 3: Incompatible monotonic choice probabilities

$$\begin{array}{llll}
P_1(v_1) & = 0 = & A_{111} & (a) \\
P_2(v_1) & = 0 = & A_{211} + A_{213} + A_{222} + A_{223} + A_{221} & (b) \\
P_3(v_1) & = 1 = & A_{311} + A_{313} + A_{333} + A_{323} & (c) \\
P_1(v_2) & = 0 = & A_{111} + A_{311} + A_{313} + A_{213} + A_{211} & (d) \\
P_2(v_2) & = 1 = & A_{221} + A_{223} + A_{323} + A_{222} & (e) \\
P_3(v_2) & = 0 = & A_{333} & (f) \\
P_1(v_3) & = 1 = & A_{111} + A_{311} + A_{211} + A_{221} & (g) \\
P_2(v_3) & = 0 = & A_{222} & (h) \\
P_3(v_1) & = 0 = & A_{313} + A_{333} + A_{213} + A_{223} + A_{323} & (i)
\end{array}$$

does not have a non-negative solution for A. To see this, note that equations (a), (b) and (d) imply $A_{111} = A_{211} = A_{213} = A_{222} = A_{223} = A_{221} = A_{333} = A_{311} = A_{313} = 0$. The values do not satisfy equation (g).

5 An Example

This section illustrates the necessary and sufficient conditions of theorem 2. The choice probabilities at the observed utility components v_t, $t = 1, \ldots, T$ are given by the Nested Multinomial Logit (NML) model of McFadden (1978) (see also Maddala (1983), pp. 67-69). We consider an NML model with three alternatives ($I = 3$). The joint distribution of the random components of the utilities is

$$F(\varepsilon) = \exp\left\{-\left[\exp(-\varepsilon_1/\theta) + \exp(-\varepsilon_2/\theta)\right]^\theta - \exp(-\varepsilon_3)\right\}.$$

The stochastic components of alternatives 1 and 2 are correlated. The association parameter θ determines the strength of the correlation: if $\theta = 1$, then ε_1 and ε_2 are stochastically independent, if $\theta \downarrow 0$ the joint distribution function of $(\varepsilon_1, \varepsilon_2)$ converges to

$$F(\varepsilon_1, \varepsilon_2) = \exp\left\{-\exp\left[-\min(\varepsilon_1, \varepsilon_2)\right]\right\}.$$

This joint distribution implies that if *e.g.* $v_1 > v_2$, then alternative 1 is eliminated from the choice set. Note that ε_3 is independent of ε_1 and ε_2.

If we take the first alternative as the reference alternative, the distribution function of $w \equiv (\varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_1)'$ becomes

$$H_1(w) = \frac{(1 + \exp(-w_2/\theta))^{(\theta-1)}}{\exp(-w_1) + (1 + \exp(-w_2/\theta))^\theta}. \tag{38}$$

The corresponding density function is

$$h_1(w_1, w_2) = \exp(-w_2)\exp(-w_1/\theta)$$

$$\times \frac{(1+\exp(-w_1))^{(\theta-2)}}{\exp(-w_2) + (1+\exp(-w_1/\theta))^{\theta}}$$

$$\times \left\{ \frac{2\,(1+\exp(-w_1/\theta))^{\theta}}{\exp(-w_2) + (1+\exp(-w_1/\theta))^{\theta}} - \frac{\theta-1}{\theta} \right\}$$

$$\times \frac{1}{\exp(-w_2) + (1+\exp(-w_1/\theta))^{\theta}}.$$

(*cf.* Börsch-Supan (1990), equations (13) and (14) which are not correct). This density is signed by the term in braces. It is clear that $0 < \theta \le 1$ is both a necessary and sufficient condition for h_1 to be non-negative on R^2.

If $\theta > 1$, there exists a set with positive measure in $\mathbb{R}^2$ where $h_1(w)$ is negative. Hence, the choice probabilities do not satisfy the non-negativity condition in (C1), and therefore they are not globally compatible with stochastic utility maximization. This is illustrated in figure 4. There and in the sequel we take $\theta = 2$. Suppose now we have a sample of three points ($T = 3$): $w_1 = (-1, 1)$, $w_2 = (2, 2)$ and $w_3 = (4, -2)$. The function $h_1(w)$ is negative in $w_3 = (4, -2)$.

Using equation (38), we can calculate the choice probabilities as

$$\begin{aligned} P(w_1) &= (0.36, 0.59, 0.05)' \\ P(w_2) &= (0.68, 0.25, 0.07)' \\ P(w_3) &= (0.13, 0.02, 0.85)'. \end{aligned}$$

After inspection of these choice probabilities (and figure 4), it is seen that they satisfy the necessary condition of corollary 3. Moreover, using the notation of the preceding section, we see that the choice probabilities also satisfy the necessary and sufficient condition of theorem 3: a non-negative solution for the A's is $A_{111} = 0.13, A_{113} = 0.23, A_{211} = 0, A_{213} = 0.32, A_{221} = 0, A_{222} = 0.02, A_{223} = 0.23, A_{233} = 0.02, A_{313} = 0$ and $A_{333} = 0.05$. Hence we conclude that the observations are compatible with stochastic utility maximization, even though the pseudo-density $h(w)$ is negative in w_3. Note that

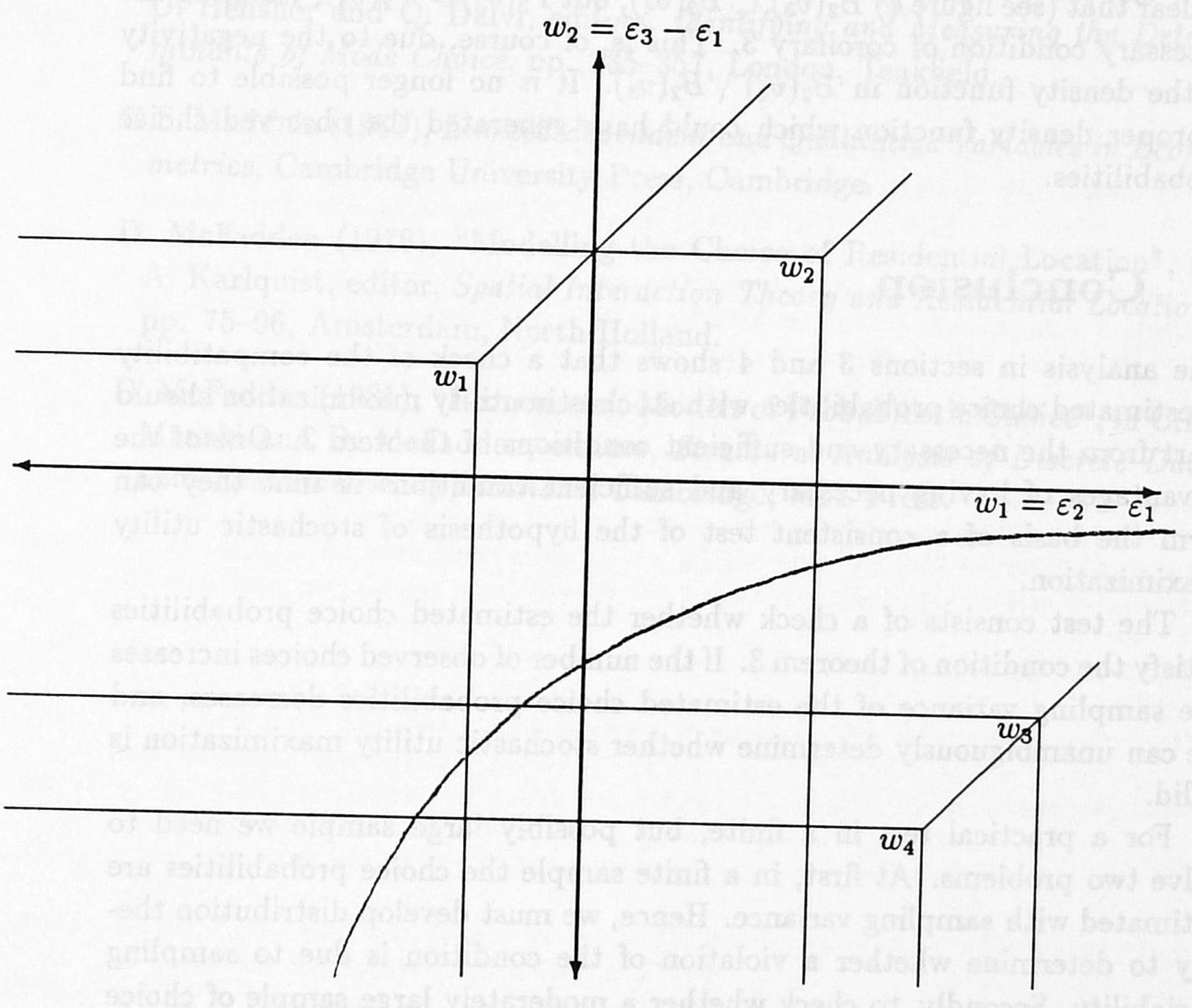

Figure 4: Choice probabilities of the nested multinomial logit model

in this case any interval that contains w_1,w_2 and w_3 violates the sufficient condition of theorem 2, and hence would lead to the conclusion that the choice probabilities are *not* compatible with stochastic utility maximization.

Now suppose we had another observation, say $w_4 = (3, -3)$. This observation has choice probabilities $(0.06, 0.01, 0.93)'$ according to the NMLM. It is clear that (see figure 4) $B_2(v_3) \subset B_2(v_4)$, but $P_2(v_3) > P_2(v_4)$, violating the necessary condition of corollary 3. This is, of course, due to the negativity of the density function in $B_2(v_4) \setminus B_2(v_3)$. It is no longer possible to find a proper density function which could have generated the observed choice probabilities.

6 Conclusion

The analysis in sections 3 and 4 shows that a check of the compatibility of estimated choice probabilities with stochastic utility maximization should start from the necessary and sufficient conditions of theorem 3. One of the advantages of having necessary and sufficient conditions is that they can form the basis of a consistent test of the hypothesis of stochastic utility maximization.

The test consists of a check whether the estimated choice probabilities satisfy the condition of theorem 3. If the number of observed choices increases the sampling variance of the estimated choice probabilities decreases, and we can unambiguously determine whether stochastic utility maximization is valid.

For a practical test in a finite, but possibly large sample we need to solve two problems. At first, in a finite sample the choice probabilities are estimated with sampling variance. Hence, we must develop distribution theory to determine whether a violation of the condition is due to sampling variability. Secondly, to check whether a moderately large sample of choice probabilities is compatible with stochastic utility maximization we need an efficient algorithm to determine whether the equation system of theorem 3 has a non-negative solution.

References

A. Börsch-Supan (1990), "On the Compatibility of Nested Logit Models with Utility Maximization", *Journal of Econometrics*, 43, pp. 371-387.

A. Daly and S. Zachary (1979), "Improved Multiple Choice Models", in D. Hensher and Q. Dalvi, editors, *Identifying and Measuring the Determinants of Mode Choice*, pp. 335–357, London, Teakfield.

G.S. Maddala (1983), *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

D. McFadden (1978), "Modelling the Choice of Residential Location", in A. Karlquist, editor, *Spatial Interaction Theory and Residential Location*, pp. 75–96, Amsterdam, North-Holland.

D. McFadden (1981), "Econometric Models of Probabilistic Choice", in C.F. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MIT Press.

Papers issued in these series from 1990 onwards: [1])

348. Pierre van der Eng, Growth and productivity change in Indonesian agriculture, 1880-1985 (1990).
349. I.P. van den Berg, Exponential stability and the river-phenomenon (1990).
350. Charles van Marrewijk, An allocation mechanism, a measure of risk aversion and an experiment (1990).
351. Ruud H. Koning, Heinz Neudecker, Tom Wansbeek, Block Kronecker products and the vecb operator (1990).
352. Peter S.H. Leeflang and Dick R. Wittink, Diagnosing competitive reactions using (aggregated) scanner data (1990).
353. Barbara L. Wolfe, Cees G.M. Sterks, Robert H. Haveman, Glen G. Cain, Estimating employment effects of tax and transfer policies: a review of simulation methods (1990).
354. A. Szirmai and D. Pilat, The international comparison of real output and labour productivity in manufacturing: a study for Japan, South Korea and the USA for 1975 (1990).
355. W. Voorhoeve, H.W.A. Dietzenbacher, N. Heerink, De GRECON-Databank. Beschrijving van de tijdreeksen die ten grondslag liggen aan de GRECON-modellen; ontleend aan officiële (CBS) publicaties (1990).
356. S.K. Kuipers en E. Sterken, Over de reikwijdte van het Nederlandse monetaire beleid (1990).
357. I.P. van den Berg, Macroscopes, regular variation and polynomial growth (1990).
358. S. Sánchez-Pedreño G., Équations différentielles hautement non linéaires ou perturbations singulières exponentielles (1990).
359. Jack A.A. van der Veen, An o(n) algorithm to solve the bottleneck traveling salesman problem restricted to ordered product matrices (1990).
360. Bas Bakker, A mundell-fleming model with a commercial banking sector (1990).
361. Maarten C.W. Janssen and Herman Nieuwenhuis, Bertrand competition and Keynesian economics (1990).
362. G.J. van den Berg, The effect of an increase of the rate of arrival of job offers on the duration of unemployment (1990).
363. Charles van Marrewijk, Endogenous population growth and the economy (1990).
364. Jack A.A. van der Veen, A new class of pyramidal solvable symmetric traveling salesman problems (1990).
365. Gerard Sierksma and Gert A. Tijssen, Asymmetry on the symmetric traveling salesman polytope (1990).
366. Gert A. Tijssen and Gerard Sierksma, On the diameter of the traveling salesman polytope (1990).
367. Charles van Marrewijk, Endogenous trade uncertainty and specialization against comparative advantage (1990).
368. Aerdt Houben, An international comparison of real output, labour productivity and purchasing power in the mineral industries of the United States, Brazil and Mexico for 1975 (1990).
369. J.C. Frowein, Modellen voor specificatie van expertise (1990).

---

[1]) These research memoranda can be ordered at the Institute of Economic Research, University of Groningen, P.O. Box 800, 9700 AV Groningen.

Memoranda marked * are out of print!

370. Lambert Schoonbeek, Stackelberg price leadership in the linear hetereogeneous duopoly (1990).

371.* Hans van Ees and Harry Garretsen, The European merger wave and 1992: self-fulfilling or self-defeating prophecies? (1990).

372. Nanne Brunia, Harry Garretsen and Robert Lensink, International spillover effects and wage flexibility: The case of Europe (1990).

373. Robert Lensink and Niels Hermes, Financial Requirements of the least developed countries through the 1990s (1990).

374. Dirk Pilat and André Hofman, Argentina's manufacturing performance: a comparative view (1990).

375. Rainer E. Burkard and Jack A.A. van der Veen, Universal conditions for algebraic traveling salesman problems to be efficiently solvable (1990).

376. Elmer Sterken, The stability of the Dutch money demand function (1990).

377. W. Voorhoeve en H.W.A. Dietzenbacher, Het model GRECON 90-A; voorspellingen 1990 en 1991 (1990).

378. Steven Brakman and Jan Jacobs, The Dutch export equation: a cointegration approach (1990).

379. T.K. Dijkstra, On statistical inference, with parameter estimates on the boundary of the parameter space (1990).

380. Theo K. Dijkstra, Pyrrho's lemma, or have it your way (1990).

381. Niels Hermes and Robert Lensink, Capital flight from six sub-saharan African countries (1990).

382. Cars H. Hommes, Helena E. Nusse, András Simonovits, Hicksian cycles and chaos in a socialist economy (1990).

383. Gusta Renes and Geert Ridder, Are women overqualified? (1990).

384. A.L.M. Abdul-Ghafoor, Simultaneity coefficients (1990).

385. dr. Herman Jacobs, Alterations in wage-determination processes (1990).

386. dr. Herman Jacobs, Job- and wage-structures in a dynamic economy (1990).

387.* J.C. Hoekstra en P.S. Zwart, Perceptie, preferentie en kennis van direct mail en andere media in de banksector (1990).

388. Ruud H. Koning en Peter P.M. Smid, Optiewaardering. Theorie en toepassingen in de investeringsplanning (1990).

389. M.A. Allers, Kosten van overdrachten voor ondernemingen (1990).

390. Pieter W. Otter, Canonical Correlation, Wiener-Granger Causality and Multi-period Ahead Prediction in Multivariate Time Series: Theory and an Application (1990).

391. Frans M. Tempelaar, The unanimity foundation of corporate financial theory (1990).

392. Jack A.A. van der Veen and René van Dal, Solvable cases of the no-wait flow-shop scheduling problem (1990).

393. P.S. Zwart en H.G.J. Gankema, Het exportgedrag van het Midden- en Kleinbedrijf in Noord-Nederland (1990).

394. J. Oosterhaven, Input-output analyse en de dienstensector (1990).

395. Harry Garretsen, Money and Nominal Prices under Uncertainty: A Suggestion (1990).

396. W. Voorhoeve en H.W.A. Dietzenbacher, Het model GRECON 90-B; septembervoorspellingen voor 1991 (1990).

397. Rainer Fremdling and Günter Knieps, Competition, regulation and nationalization the Prussian railroad system in the 19th century (1990).

398. I.P. van den Berg, Macroscopic rivers (1990).

399. Cars H. Hommes, Adaptive expectations and nonlinear cobweb models: stable cycles, chaos and bifurcations (1990).

400. Georg Tillmann, Income equality and income taxation (1990).

401. C. Praagman and J.M. Soethoudt, Continuous time behaviors (1990).

402. Bas Bakker, A suggested solution to the Feldstein-Horioka paradox (1990).
403. H. van Ees en H. Garretsen, Sparen, investeren en de vermogensmarkt (1990).
404. Herman J. de Jong, Infrastructure and inland transport in the Netherlands during the first railway years, 1800-1880 (1991).
405. Hans van Ees, Does money affect money prices? (1991).
406. Bas Bakker, Het Netto Buitenlands Actief van de banken in de periode 84-90 (1991).
407. Claude Wetta, Marché et mobilité de l'emploi au Burkina Faso (1991).
408. Dirk Pilat, Levels of real output and labour productivity by industry of origin, a comparison of Japan and the United States, 1975 and 1970-1987 (1991).
409. Piet Boomsma, Jan Oosterhaven and Anne van der Veen, Construction of a representative Dutch bi-regional input-output table (1991).
410. H.J. Brinkman, J.W. Drukker, S.J. Stuurop, De sociale afkomst der Nederlandse lotelingen in de 19e eeuw (1991).
411. Elmer Sterken, Models of the business cycle; a survey (1991).
412. Bas Bakker, Is debt neutral when firms are cash-in-advance constrained? (1991).
413. Arend Buitendam, Decentralization and the governance of employment relationships: human resource management between labour market and organization (1991).
414. Ruud H. Koning and Geert Ridder, Discrete choice and stochastic utility maximization (1991).